

Conference Abstract

# FAIR data in meta-omics research: Using the MOD-CO schema to describe structural and operational elements of workflows from field to publication

Janno Harjes<sup>‡</sup>, Dagmar Triebel<sup>§</sup>, Anton Link<sup>§</sup>, Tanja Weibulat<sup>‡</sup>, Frank Oliver Glöckner<sup>¶</sup>, Gerhard Rambold<sup>‡</sup>

<sup>‡</sup> University of Bayreuth, Bayreuth, Germany

<sup>§</sup> Staatliche Naturwissenschaftliche Sammlungen Bayerns, Munich, Germany

<sup>‡</sup> German Federation for Biological Data e.V., Bremen, Germany

<sup>¶</sup> Alfred-Wegener-Institut, Helmholtz Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany

Corresponding author: Janno Harjes ([janno7806@gmail.com](mailto:janno7806@gmail.com))

Received: 21 Jun 2019 | Published: 02 Jul 2019

Citation: Harjes J, Triebel D, Link A, Weibulat T, Glöckner F, Rambold G (2019) FAIR data in meta-omics research: Using the MOD-CO schema to describe structural and operational elements of workflows from field to publication. Biodiversity Information Science and Standards 3: e37596. <https://doi.org/10.3897/biss.3.37596>

## Abstract

Nucleic acid and protein sequencing-based analyses are increasingly applied to determine origin, identity and traits of environmental (biological) objects and organisms. In this context, the need for corresponding data structures has become evident. As existing schemas and community standards in the domains of biodiversity and molecular biological research are comparatively limited with regard to the number of generic and specific elements, previous schemas for describing the physical and digital objects need to be replaced or expanded by new elements for covering the requirements from meta-omics techniques and operational details. On the one hand, schemas and standards are hitherto mostly focussed on elements, descriptors, or concepts that are relevant for data exchange and publication, on the other hand, detailed operational aspects regarding origin context and laboratory processing, as well as data management details, like the documentation of physical and digital object identifiers, are rather neglected.

The conceptual schema for Meta-omics Data and Collection Objects (MOD-CO; <https://www.mod-co.net/>) has been set up recently Rambold et al. 2019. It includes design

elements (descriptors or concepts), describing structural and operational details along the work- and dataflow from gathering environmental samples to the various transformation, transaction, and measurement steps in the laboratory up to sample and data publication and archiving. The concepts are named according to a multipartite naming structure, describing internal hierarchies and are arranged in concept (sub-)collections. By supporting various kinds of data record relationships, the schema allows for the concatenation of individual records of the operational segments along a workflow (Fig. 1). Thus, it may serve as a logical and structural backbone for laboratory information management systems. The concept structure in version 1.0 comprises 653 descriptors (concepts) and 1,810 predefined descriptor states, organised in 37 concept (sub-)collections. The published version 1.0 is available as various schema representations of identical content ([https://www.mod-co.net/wiki/Schema\\_Representations](https://www.mod-co.net/wiki/Schema_Representations)). A normative XSD (= XML Schema Definition) for the schema version 1.0 is available under [http://schema.mod-o.net/MOD-CO\\_1.0.xsd](http://schema.mod-o.net/MOD-CO_1.0.xsd).

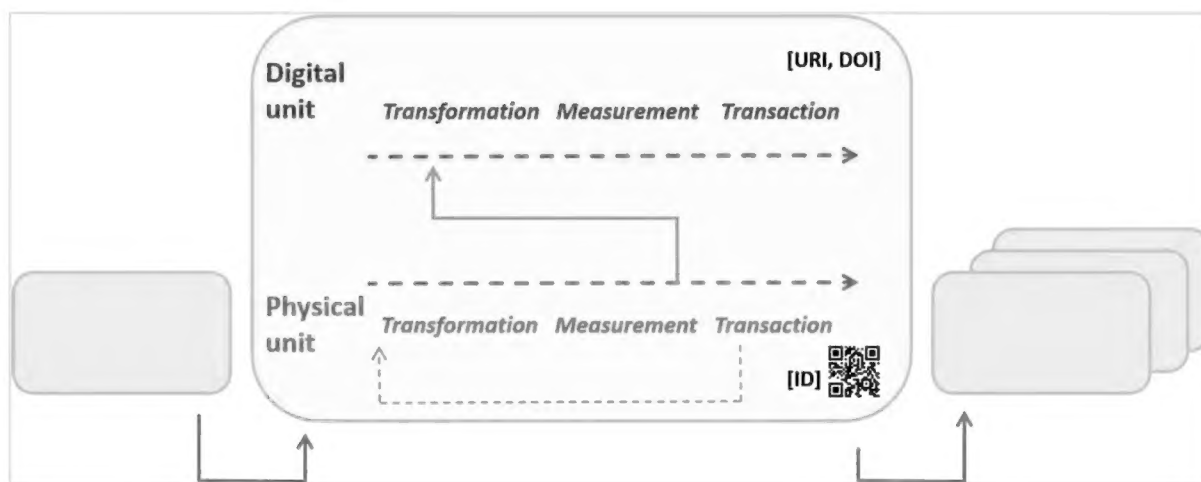


Figure 1.

Workflow segments concatenated to a single workflow. A workflow segment comprises the elementary operations transformation, measurement and transaction being applied once or twice (due to preceding subsampling) to a physical object in focus, and the generation of data (measurement) and its subsequent transformation, measurement and transaction within the segment.

The MOD-CO concepts might be integrated as descriptor/element structures in the relational database DiversityDescriptions (DWB-DD) an open-source and freely available software of the Diversity Workbench (DWB; <https://diversityworkbench.net/Portal/DiversityDescriptions>; <https://diversityworkbench.net>). Currently, DWB-DD is installed at the Data Center of the Bavarian Natural History Collections (SNSB) to build an instance of its own for storing and maintaining MOD-CO-structured meta-omics research data packages and enrich them with 'metadata' elements from the community standards Ecological Markup Language (EML), Minimum Information about any (x) Sequence (MIxS), Darwin Core (DwC) and Access to Biological Collection Data (ABCD). These activities are achieved in the context of ongoing FAIR ('Findable, Accessible, Interoperable and Reuseable') biodiversity research data publishing via the German Federation for Biological Data (GFBio) network (<https://www.gfbio.org/>). Version 1.1 of the schema with extended collections of structural and operational design concepts is scheduled for 2020.

## Keywords

collection data, conceptual schema, DiversityDescriptions, LIMS, meta-omics, workflow segments

## Presenting author

Janno Harjes

## Presented at

Biodiversity\_Next 2019

## Acknowledgements

We are grateful for discussions with Andreas Brachmann (Munich), Gregor Hagedorn (Berlin), Derek Peršoh (Bochum), Veronica Sanz (Munich), Carola Söhngen (Braunschweig), Thorsten Stoeck (Kaiserslautern), Christoph Tebbe (Braunschweig), and Pelin Yilmaz (Bremen).

## References

- Rambold G, Yilmaz P, Harjes J, Klaster S, Sanz V, Link A, Glöckner FO, Triebel D (2019) Meta-omics data and collection objects (MOD-CO): a conceptual schema and data model for processing sample data in meta-omics research. Database 2019 <https://doi.org/10.1093/database/baz002>